Received: 10 October 2019 | Revised: 8 November 2019 | Accepted: 13 November 2019

DOI: 10.1002/chir.23158

SPECIAL ISSUE ARTICLE

WILEY

A molecular orbital selection approach for fast calculations of specific rotation with density functional theory

Tal Aharon | **Marco Caricato**

Department of Chemistry, University of Kansas, Lawrence, Kansas, USA

Correspondence
Marco Caricato, Department of Chemistry, University of Kansas, 1567 Irving Hill Road, Lawrence, KS 66045.
Email: mcaricato@ku.edu

Funding information
Division of Chemistry, Grant/Award Number: CHE-1650942

Abstract

In this work, we describe a simple approach to select the most important molecular orbitals (MOs) to compute the optical rotation tensor through linear response (LR) Kohn-Sham density functional theory (KS-DFT). Taking advantage of the iterative nature of the algorithms commonly used to solve the LR equations, we select the MOs with contributions to the guess perturbed density that are larger than a certain threshold and solve the LR equations with the selected MOs only. We propose two criteria for the selection, and two definitions of the selection threshold. We then test the approach with two functionals (B3LYP and CAM-B3LYP) and two basis sets (aug-cc-pVDZ and aug-cc-pVTZ) on a set of 51 organic molecules with specific rotation spanning five orders of magnitude, 10^0–10^4 deg $(dm^{-1}\ (g/mL)^{-1})$. We show that this approach indeed can provide very accurate values of specific rotation with estimated speedup that ranges from 2 to 8× with the most conservative selection criterion, and up to 20 to 30× with the intermediate criterion.

KEYWORDS

DFT, linear response, quantum chemistry, specific rotation, speedup

1 | INTRODUCTION

The study of optically active molecules is of great interest because of the fundamental role they play as the building blocks of life (L-amino acids and D-sugars) and their importance in the pharmaceutical industry.[1] Facile assignment of the absolute configuration (AC) of chiral molecules is important because it is directly tied to their bioactive properties. The ability to directly link a structure to the chiroptical property of interest, such as optical rotation (OR), led electronic structure calculations to become a standard for identifying the AC of chiral molecules. Accurate calculations require the proper account of electron correlation and large basis sets.[1] To this end, great progress has been made in recent years in applying both density functional theory (DFT) and coupled cluster (CC) methods to the calculation of $[\alpha]_\omega$.[2-17] However, as systems of interest become larger, the cost of electronic structure calculations becomes prohibitive, and compromises between cost and accuracy are needed.[18]

Despite the increasing complexity of systems of interest, few attempts have been made to reduce the cost of $[\alpha]_\omega$ calculations as a whole. Wiberg et al calculated $[\alpha]_\omega$ with small basis sets (STO-3G and 3-21G) augmented with diffuse functions from large correlation-consistent basis sets.[19] These mixed basis sets reproduce the values obtained with the full basis sets at a fraction of the computational cost. Another example is the OR prediction (ORP) basis set, developed specifically to calculate $[\alpha]_\omega$, which is able to produce results of aug-cc-pVTZ quality but is similar in size to aug-cc-pVDZ.[20,21] Crawford et al

[This article is part of the Special Issue: In honor and memory of Prof. Koji Nakanishi. See the first articles for this special issue previously published in Volumes 31:12. More special articles will be found in this issue as well as in those to come.]

extended the local correlation idea of Pulay and Saebø[22,23] to response properties[24] and then later implemented it for $[\alpha]_\omega$ calculations.[25] This method uses localized orbitals and neglects interactions between distant orbitals, thus reducing the cost of the correlation part of the calculation.[26,27] While this method is successful for smaller systems, the authors show that for larger systems the need for tighter thresholds outweighs the benefits of neglecting parts of the wave function.[25]

In this work, we present a different approach to reduce the cost of $[\alpha]_\omega$ calculations with Kohn-Sham DFT (KS-DFT), based on the selection of the molecular orbitals (MOs) that are likely to contribute the most to this property and discarding the rest. We have recently shown that a significant portion of the MOs do not contribute significantly to $[\alpha]_\omega$ through a post-calculation analysis of the OR tensor.[28-30] Here, we try to determine the meaningful MOs beforehand and to solve the linear response equations only with a subset of relevant MOs while preserving accuracy.

The paper is organized as follows: Section 2 outlines the procedure for the orbital selection; Section 3 contains the results of the calculations; and Section 4 presents a discussion and concluding remarks.

2 | THEORY AND COMPUTATIONAL DETAILS

In order to explain how the orbital selection process works, it is useful to briefly review how the specific rotation is typically computed. For molecules in isotropic media, the specific rotation is related to trace of the electric dipole-magnetic dipole polarizability[31] tensor $G'_{\alpha\beta}$:

$$\beta(\omega) = -\frac{1}{3\omega}\sum \mathrm{G}'_{\alpha\alpha}, \tag{1}$$

where ω is the frequency of incident light, and the tensor $G'_{\alpha\beta}$ is defined as follows:

$$\mathrm{G}'_{\alpha\beta} = -2\omega\sum_{n\neq 0}\mathrm{Im}\frac{\langle\varphi_0|\mu_\alpha|\varphi_n\rangle\langle\varphi_n|m_\beta|\varphi_0\rangle}{\omega_{n0}^2-\omega^2}, \tag{2}$$

where μ is the electric dipole moment operator, m is the magnetic dipole moment operator, the Greek letter indexes represent Cartesian coordinates, φ_0 and φ_n are the ground and excited electronic state wave functions, respectively, and ω_{n0} is the corresponding transition energy. However, this series is slowly converging, and evaluating electronic excited states is computationally expensive. A significantly more efficient approach is based on time-averaged variational linear response theory,[16,32,33] where the evaluation of the transition moments in Equation (2) is replaced by the evaluation of the perturbed electron density, P^x_α, where x is the external field perturbation. Using Hartree-Fock (HF) or KS-DFT as approximations of the unperturbed density, and using the variational nature of these methods, the perturbed density can be obtained by solving the coupled perturbed HF or KS equations (CPHF or CPKS):[16,32,33]

$$\left[\begin{pmatrix}\mathbf{M} & \mathbf{Q}\\ \mathbf{Q}^* & \mathbf{M}^*\end{pmatrix}-\omega\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\right]\begin{pmatrix}\mathbf{P}^S_\beta\\ \left(\mathbf{P}^S_\beta\right)^*\end{pmatrix}=\begin{pmatrix}\mathbf{S}_\beta\\ \mathbf{S}^*_\beta\end{pmatrix}, \tag{3}$$

where $M_{ia,jb} = (\varepsilon_a - \varepsilon_i)\delta_{ij}\delta_{ab}+K_{aj,ib}$, $Q_{ia,jb} = K_{ab,ij}$, $K_{ia,jb} = \langle aj|\, ib\rangle+\langle aj|\, f^{XC}(r,r')|\, ib\rangle$, and $\mathbf{1}$ is the unit matrix. Extra terms should appear in Equation (3) when perturbation-dependent atomic orbitals (such as gauge including atomic orbitals, or GIAOs[16,34,35]) are used, but they are neglected here for simplicity. The indices i and j denote occupied MOs, a and b virtual MOs, ε is the MO energies, $\langle aj|\, ib\rangle$ are two electron repulsion integrals (2ERIs), and $f^{XC}(r,r')$ is the exchange correlation kernel of the density functional. In Equation (3), the matrices $S_{\beta,ai} = \langle a|\, S_\beta|\, i\rangle$ and their complex conjugate represent the perturbation integrals in MO basis: $S_\beta = \mu_\beta$ or m_β. The G' tensor is then obtained by tracing the perturbed density with the conjugate dipole integrals:

$$\mathrm{G}'_{\alpha\beta} = -\omega\mathrm{Tr}\left(\mathrm{Im}\left\{\mathbf{R}_\alpha\left(\mathbf{P}^S_\beta\right)^*\right\}\right), \tag{4}$$

where $\mathbf{R}_\alpha = \mathbf{m}_\alpha$ or $\boldsymbol{\mu}_\alpha$.

Equation (3) is generally not solved by direct inversion of the matrix on the left-hand side because the transformation of the 2ERIs to MO basis is computationally expensive (it scales as $O(N^5)$, where N is the basis set size), and the matrix itself is rather large (ie, the dimension is $N_{\mathrm{occ}} * N_{\mathrm{vir}}$, where N_{occ} is the number of occupied MOs and N_{vir} is the number of virtual MOs). Instead, efficient iterative algorithms based on partial AO $\leftrightarrow$ MO basis transformations are employed.[36,37] These algorithms avoid the full 2ERIs transformation and the storage of the $\mathbf{M}$ and $\mathbf{Q}$ matrices. The scaling is thus reduced formally to $O(N^4)$ but practically to $O(N^3)$ by using efficient integral evaluation and contraction algorithms.

In the context of this work, the key point is that any iterative algorithm generates a guess for the solution vectors; see eq. (64) in Pople et al.,[36] to start the solution of Equation (3). Thus, we can use the guess $\mathbf{P}^S$ to define two possible criteria for the MO selection, one based on the square elements of the guess matrix:

$$G^1{}_{ia} \equiv \sum_{\beta} \left| P^{S}_{\beta,ia} \right|^2, \tag{5}$$

and one based on the dot product with the conjugate dipole integrals:

$$G^2{}_{ia} \equiv \tilde{S}_{ia} = \mathrm{Im}\left\{ \sum_{\beta} P^{\boldsymbol{S}}_{\beta,ai} \left(R_{\beta,ai} \right)^* \right\}, \tag{6}$$

where we have used the $\tilde{S}$ symbol to connect this quantity with the rotatory strength in configuration space, which we defined for the qualitative analysis of the MO transition contributions to the OR tensor.[28-30] The difference here is that the $\tilde{S}$ values are computed with the guess density rather than the converged density. In particular, we solve the CPKS equations for the magnetic dipole perturbation, so that $S = m$ and $R = \mu$ in Equations (5) and (6). Therefore, an occupied MO is selected if

$$\left| G^1{}_{ia} \right| \text{or} \left| G^2{}_{ia} \right| > \varepsilon \quad \forall a, \tag{7}$$

and a virtual MO is selected if

$$\left| G^1{}_{ia} \right| \text{or} \left| G^2{}_{ia} \right| > \varepsilon \quad \forall i, \tag{8}$$

depending on the choice of selection criterion. The threshold ε is defined either as an absolute value, e.g., $\varepsilon = 10^{-n}$ where n is a positive integer or relative to the largest value of $|G_{ia}|$, e.g., $\varepsilon = 10^{-n}$ Max$\{|G_{ia}|\}$. In the following section, we test these four possibilities for the MO selection criterion with three values of n: $n = 3, 4, 5$.

Since the scope of this work is only to test the validity of this idea, we remove the discarded MOs from the calculation simply by setting to zero the corresponding unperturbed MO coefficients in the CPKS calculation. This ensures that these MOs do not contribute to the evaluation of the OR tensor, but it does not effectively change the cost of the calculation. Instead, we estimate the potential cost savings by assuming a cubic scaling with the size of the basis set, as mentioned above, and reporting a speedup defined as follows:

$$\text{Speed} - \text{Up} = \left(\frac{N_{tot}}{N_{sel}} \right)^3, \tag{9}$$

where N_{tot} is the total number of MOs and N_{sel} is the number of selected MOs. All calculations were performed with a development version of the GAUSSIAN suite of programs.[38] Geometries were optimized using the CAM-B3LYP/aug-cc-pVDZ model chemistry.[39,40] Calculations of specific rotation were performed at the sodium D-line (ω = 589.3 nm) with the B3LYP[41-43] and CAM-B3LYP functionals and the aug-cc-pVDZ and aug-cc-pVTZ basis sets in the length gauge formalism using GIAOs.[16,34,35]

3 | RESULTS

To test the success of the selection procedure outlined in Section 2, we performed calculations using a total of 51 molecules. This set is constructed from the 42 organic molecules provided in the OR45 test set of Srebro et al,[18] plus seven more organic molecules shown in Figure 1: two conformers of 2-carene (**43** and **44**), two conformers of 3-methylcyclopentanone (**45** and **46**), cycloserine (**47**), fucose (**48**), limonene (**49**), nicotine (**50**), and oxaceprol (**51**). The absolute values of the specific rotation range across five orders of magnitude, 10^0–10^4 deg (dm^{-1}

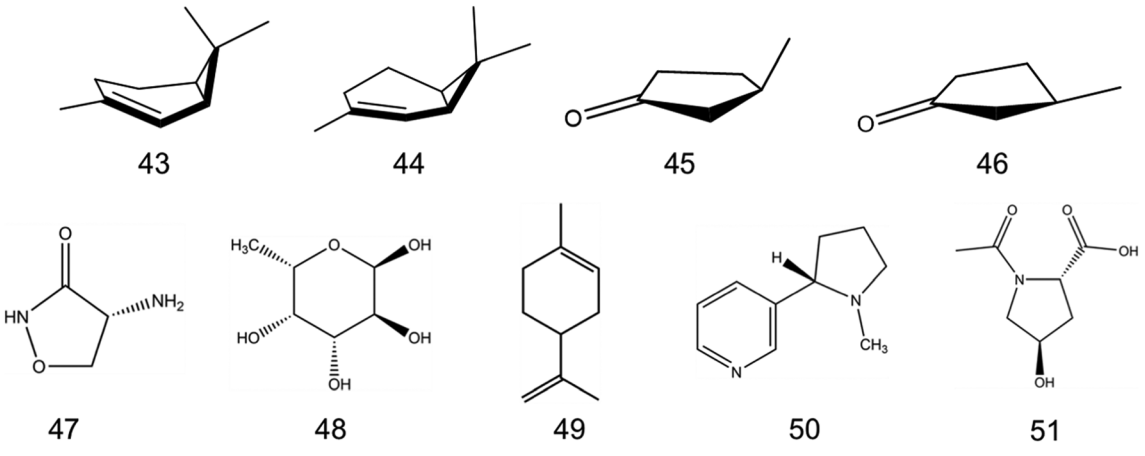

FIGURE 1 Structures of molecules **43-51**

$(g/mL)^{-1}$), reported in Tables S1 to S8 in the Supporting Information (SI), providing a stringent test for the procedure.

We define a short-hand notation to distinguish the choice of selection criterion: A/R represents absolute or relative thresholds; G/S represents G^1 or G^2 (in Equations (5) and (6)), respectively, followed by the threshold order of magnitude (n). For instance, AG5 indicates the G^1 criterion with the absolute threshold $\varepsilon = 10^{-5}$. The results are presented in terms of relative errors (%), with histograms that collect the error distribution for the test set and tables that include the mean signed error (MSE), mean unsigned error (MUE), maximum error (Max), the MSE standard deviation (σ_{MSE}), the average relative number of selected MOs ($\tilde{N}_{sel}$), its standard deviation ($\sigma_{\tilde{N}_{sel}}$), and the estimated speedup based on Equation (9). All values of $[\alpha]_D$ with all selection criteria are reported in Tables S1 to S8 of the SI, while Tables S9 to S16 report the number of occupied and virtual MOs employed in each calculation.

The relative error distribution for CAM-B3LYP/aug-cc-pVDZ and the AGn criterion is reported in Figure 2. The plots for the other three selection criteria are similar and are reported in Figures S1 to S3. The collective statistical analysis for this model chemistry and all selection criteria is reported in Table 1. It is immediately apparent from the figure that the error distribution for AG5 is sharply centered around 0, and it spreads out as n decreases. Similar trends are obtained with the other criteria. All calculations with thresholds $n > 3$ reproduce $[\alpha]_D$ very well, as indicated by both the small MUE and σ_{MSE} in Table 1. The AG5 results are very accurate, with a MUE of only 0.8%, a σ_{MSE} of 1.4%, and a Max of only 5.5%. This selection criterion discards about 40% of the MOs, with an estimated speedup of 4×, which is already significant. Using $n = 4$ leads to a slight drop in accuracy, but MSE and MUE are still small, −1.1% and 3.9%, respectively. The values of σ_{MSE} and Max are larger, 8.5% and 49%, respectively. However, this is due to molecules with small absolute values of $[\alpha]_D$, where small absolute errors may result in seemingly large relative errors. For instance, the Max = 49% comes from molecule **1**, which has $[\alpha]_D = -19$ deg $(dm^{-1}\ (g/mL)^{-1})$, and an absolute error of −9 deg $(dm^{-1}\ (g/mL)^{-1})$; note that both the $[\alpha]_D$ value and the error are considerably below the average expected error for this choice of model chemistry, ie, 25–30 deg $(dm^{-1}\ (g/mL)^{-1})$.[18] The other two molecules with seemingly large relative error in Figure 2 for AG4 are **19**, which has $[\alpha]_D = 10$ deg $(dm^{-1}\ (g/mL)^{-1})$ and an absolute error of 2.1 deg $(dm^{-1}\ (g/mL)^{-1})$, and **6**, with $[\alpha]_D = 15$ deg $(dm^{-1}\ (g/mL)^{-1})$ and an absolute error of 2.4 deg $(dm^{-1}\ (g/mL)^{-1})$. Therefore, the AG4 criterion does in fact provide even better results than what appears in Table 1 if the statistical analysis had been performed subdividing the test set in compounds with large and small $[\alpha]_D$, and the absolute rather than the relative error had been used for the latter set. On average, this selection criterion eliminates more than half of the

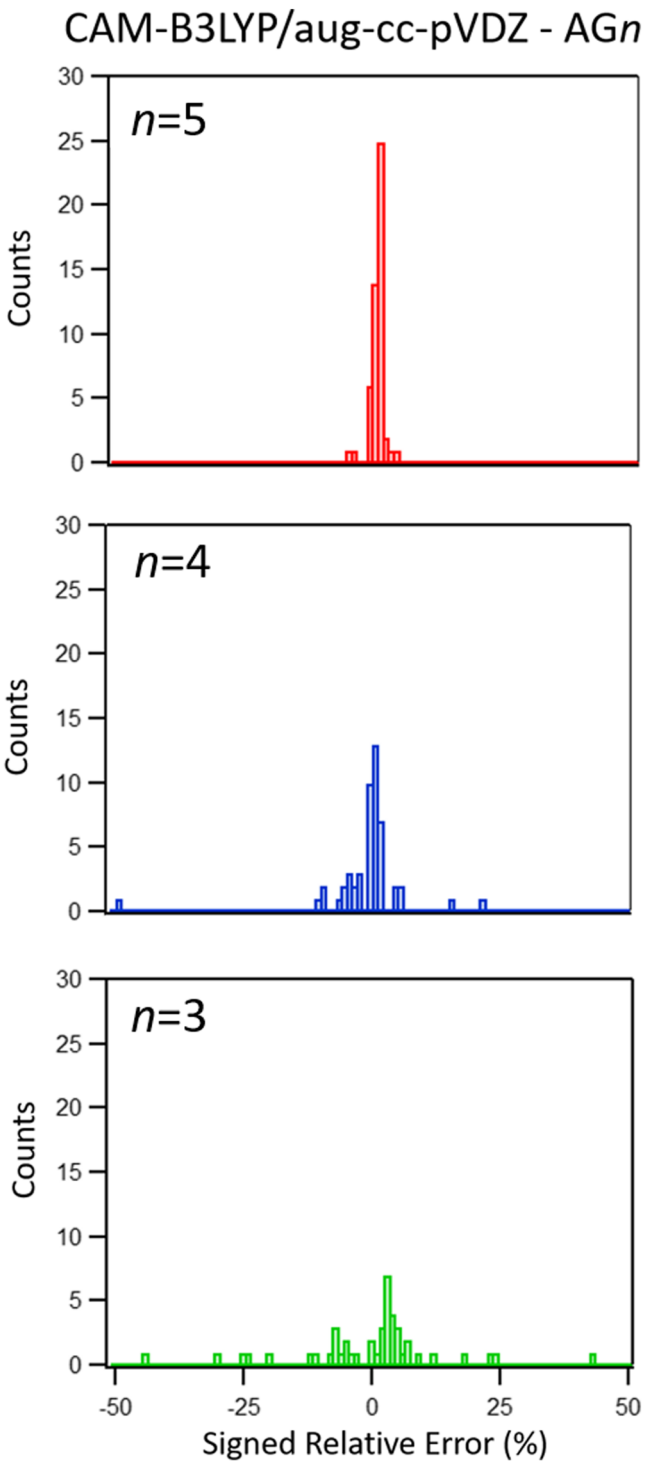

FIGURE 2 Histogram of the signed relative errors for each of the 51 molecules using the AGn selection criterion and the CAM-B3LYP/aug-cc-pVDZ method

TABLE 1 Relative (%) error: MSE, MUE, Max, and σ_{MSE}, average fraction (%) of selected MOs ($\tilde{N}_{sel}$) and the corresponding standard deviation ($\sigma_{\tilde{N}_{sel}}$), and estimate of the average speedup (S.U., see Equation (9)) for calculations with the aug-cc-pVDZ basis set

	CAM-B3LYP			B3LYP		
n	**5**	**4**	**3**	**5**	**4**	**3**
			AG*n*			
MSE	−0.1	−1.1	−1.7	−0.2	−0.3	−7.8
MUE	0.8	3.9	23	0.7	3.5	21
Max	5.5	49	211	7.3	33	228
σ_{MSE}	1.4	8.5	51	1.5	7.7	47
$\tilde{N}_{sel}$	62	45	24	63	46	25
$\sigma_{\tilde{N}_{sel}}$	5	5	4	5	5	3
S.U.	4.1	11	70	4	11	60
			RG*n*			
MSE	0	0.6	−0.7	0.1	0.1	−5.3
MUE	0.2	1.4	6.1	0.7	2.4	17
Max	1.7	13	28	9.4	21	161
σ_{MSE}	0.4	2.6	10	1.9	4.7	37
$\tilde{N}_{sel}$	72	57	37	63	47	27
$\sigma_{\tilde{N}_{sel}}$	10	12	11	15	14	12
S.U.	2.6	5.5	20	4	10	50
			AS*n*			
MSE	0	−0.6	−4.4	0	−0.1	−10
MUE	0.2	1.9	16.3	0.2	2	18
Max	1.2	9.5	170	1.5	18	200
σ_{MSE}	0.3	3.2	36	0.4	4.4	43
$\tilde{N}_{sel}$	73	54	28	72	56	29
$\sigma_{\tilde{N}_{sel}}$	10	5	5	9	5	5
S.U.	2.6	6	43	2.6	6	42
			RS*n*			
MSE	0.1	0.1	0.2	0	0.1	0.3
MUE	0.2	0.2	1.9	0.1	0.2	2
Max	1.7	1.5	9.3	1.1	0.9	11
σ_{MSE}	0.4	0.4	2.9	0.3	0.3	3.3
$\tilde{N}_{sel}$	79	71	55	75	67	49
$\sigma_{\tilde{N}_{sel}}$	11	12	12	11	13	14
S.U.	2	2.8	6.2	2.4	3.3	8.3

Abbreviations: MO, molecular orbital; MSE, mean signed error; MUE, mean unsigned error.

MOs (55%), with a considerable estimated speedup of 11×. On the other hand, the AG3 criterion is too loose, as shown both in Figure 2 and in Table 1, and the large speedup comes at the price of poor accuracy.

The error distributions for the RG*n* criterion at CAM-B3LYP/aug-cc-pVDZ level are shown in Figure S1 of the SI, and the statistical data are reported in Table 1. This criterion provides essentially the same trends as the corresponding AG*n* choice. However, the error averages and spread tend to be smaller than those for AG*n* for the same level of *n*. This is because a smaller number of MOs (about 10% less) are discarded with RG*n* than with AG*n* for every *n*. The better accuracy comes with smaller speedup by a factor of approximately 2. The results for the AS*n* criterion are reported in Figure S2 of the SI and

Table 1. Calculations performed with this criterion provide results that are very close to those with RGn in terms of accuracy, number of discarded MOs, and corresponding speedup. Finally, the RSn data, reported in Figure S3 of the SI and Table 1, are virtually identical to those obtained with the RG(n + 1) criterion. Also for these criteria, the largest relative errors are due to molecules with small $[\alpha]_D$; for instance, the Max error for RG4 (13%) is due to molecule **18**, with $[\alpha]_D$ = 2.3 deg (dm^{-1} (g/mL)$^{-1}$) and an error of 0.29 deg (dm^{-1} (g/mL)$^{-1}$).

The error distributions with the B3LYP/aug-cc-pVDZ model chemistry and the AGn criterion are shown in Figure 3, while similar plots for the other criteria are shown in Figures S4–S6 of the SI. The statistical analysis for all criteria is also reported in Table 1. The qualitative trends in the error distribution are very similar to those for the CAM-B3LYP functional. The statistical data also mirror that of CAM-B3LYP, both in terms of accuracy and computational cost. The only notable difference is that the errors with the RGn criterion are larger than those for CAM-B3LYP because a larger number of MOs are discarded. A similar behavior, albeit to a smaller extent, is obtained with the RSn criterion. This difference is due to the fact that the magnitude of the larger G_{ia}^1 (and G_{ia}^2) values is larger with CAM-B3LYP than with B3LYP, which results in a looser selection threshold for the latter method. Six molecules in the AG4 set have errors greater than 10%: **1**, **6**, **9**, **18**, **19**, and **34**, with errors of 33%, 28%, 15%, 19%, 16%, and 14%, respectively. The $[\alpha]_D$ for these molecules are −17, 12, 10, 1.9, 9, and −9 deg (dm^{-1} (g/mL)$^{-1}$), respectively, while the errors are 5.6, 3.24, 1.46, 0.35, 1.34, and 1.25 deg (dm^{-1} (g/mL)$^{-1}$), respectively. With RG4, molecules **9**, **18**, and **34** have a relative error greater than 10%: −15%, −15%, and 21%, corresponding to absolute values of 1.46, 0.27, and 1.86 deg (dm^{-1} (g/mL)$^{-1}$), respectively, for $[\alpha]_D$ values of −10, 1.9, and −9 deg (dm^{-1} (g/mL)$^{-1}$), respectively. The B3LYP AS4 set contains only three molecules with errors greater than 10% (**1**, **9**, and **18**), where the errors are 17%, 18%, and 11%, respectively, and the corresponding $[\alpha]_D$ values are also small: −17, −10, and 1.90 deg (dm^{-1} (g/mL)$^{-1}$), respectively. No errors greater than 10% are found for the RS4 criterion. These values of $[\alpha]_D$ and corresponding errors are again below the expected accuracy of this level of theory, estimated to be in the 20 to 25 deg(dm^{-1} (g/mL)$^{-1}$) range.[18,44]

The error distribution for CAM-B3LYP and B3LYP with the aug-cc-pVTZ basis set and the AGn criterion are shown in Figures 4 and 5, while the same plots with the other selection criteria are reported in Figures S7–S12 of the SI. The statistical data of the error and computational cost are shown in Table 2. The qualitative trends of error distribution with this basis set are similar to those with aug-cc-pVDZ, except that the distribution is now slightly more spread out. The statistical data in the table for CAM-B3LYP show that, for each selection criterion, there seems to be a correspondence with the n-1 choice of the aug-ccpVDZ basis set in terms of accuracy, amount of

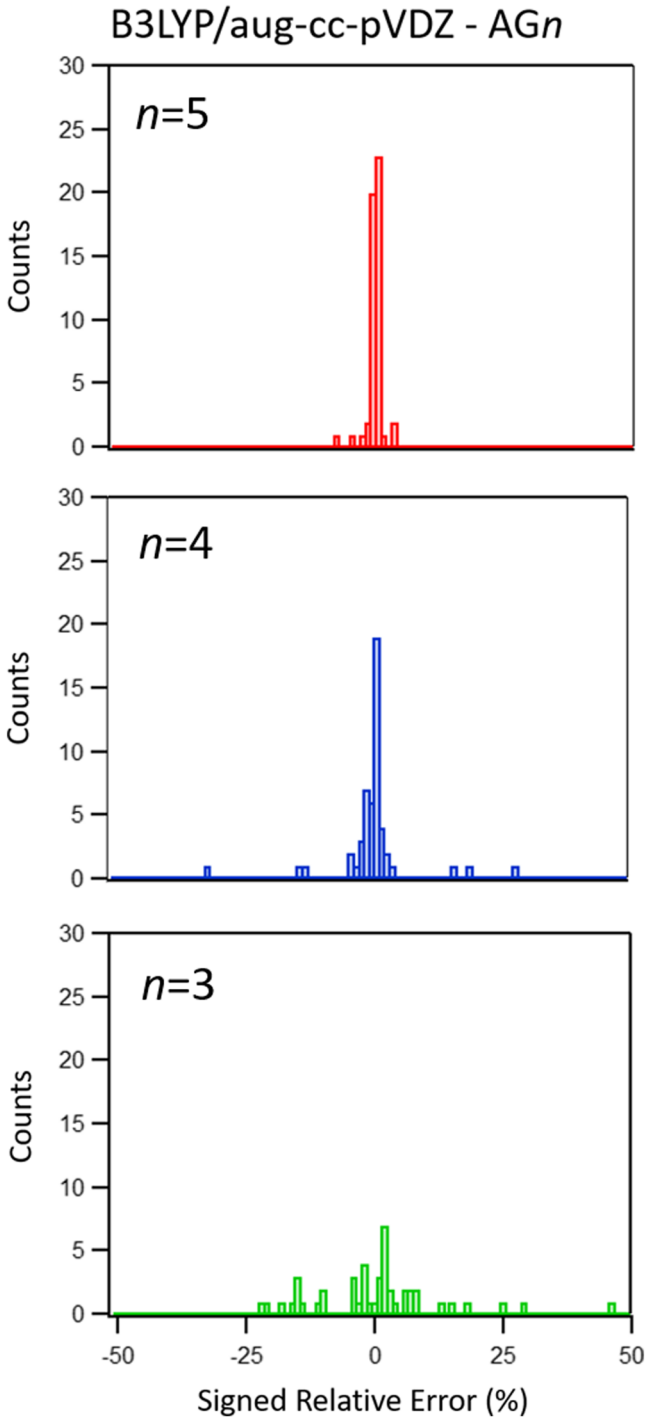

FIGURE 3 Histogram of the signed relative errors for each of the 51 molecules using the AGn selection criterion and the B3LYP/aug-cc-pVDZ method

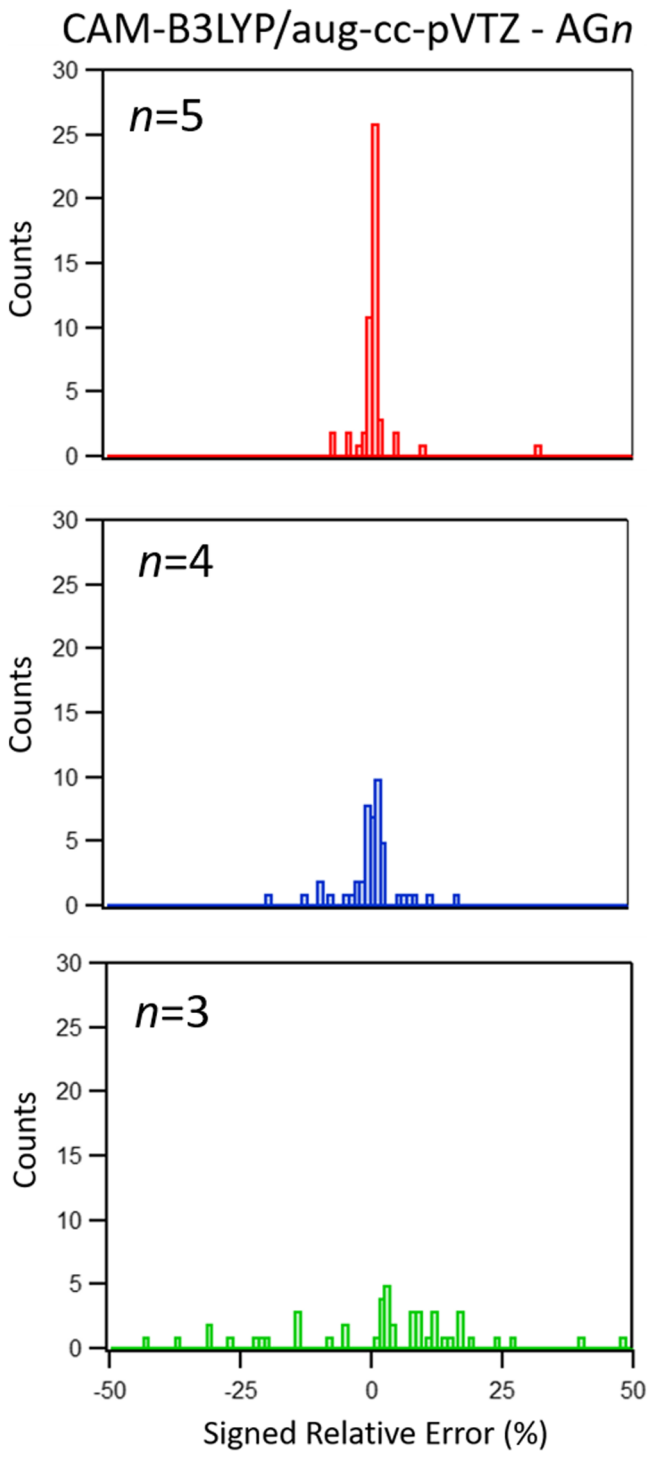

FIGURE 4 Histogram of the signed relative errors for each of the 51 molecules using the AG*n* selection criterion and the CAM-B3LYP/aug-cc-pVTZ method

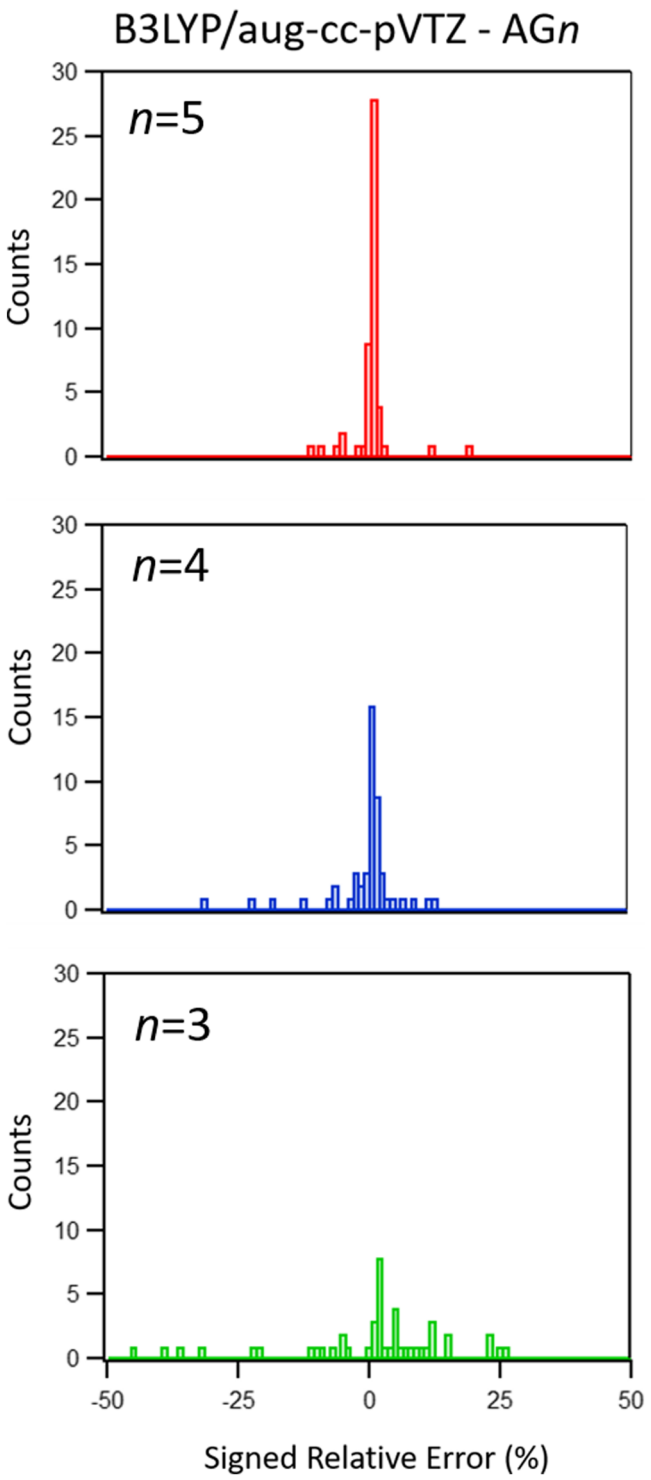

FIGURE 5 Histogram of the signed relative errors for each of the 51 molecules using the AG*n* selection criterion and the B3LYP/aug-cc-pVTZ method

selected MOs, and speedup. This is related to the fact that with a larger basis set there are more virtual MOs available, corresponding to a larger number of $P^S_{\beta,ai}$ elements. More of these elements are small compared to the aug-cc-pVDZ case, but they still contribute significantly to the final value of the specific rotation. Nonetheless, a tight choice of n ($n = 5$) for aug-cc-pVTZ provides smaller spread (both in terms of σ_{MSE} and Max) and a larger speedup than with $n = 4$ for aug-cc-pVDZ, indicating that a tighter threshold can be used for the larger basis set without sacrificing accuracy and computational gains. In

TABLE 2 Relative (%) error: MSE, MUE, Max, and σ_{MSE}, average fraction (%) of selected MOs ($\tilde{N}_{sel}$) and the corresponding standard deviation ($\sigma_{\tilde{N}_{sel}}$), and estimate of the average speedup (S.U., see Equation (9)) for calculations with the aug-cc-pVTZ basis set

	CAM-B3LYP			B3LYP		
n	**5**	**4**	**3**	**5**	**4**	**3**
			AG*n*			
MSE	0.6	4.6	−72	0.1	−2.2	−5.4
MUE	1.9	12	95	1.8	5	22
Max	31	230	3,698	19	60	199
σ_{MSE}	5.1	38	519	4.1	11	46
$\tilde{N}_{sel}$	39	23	12	39	23	13
$\sigma_{\tilde{N}_{sel}}$	3	3	2	3	3	2
S.U.	17	85	600	17	79	500
			RG*n*			
MSE	0.3	1	2.7	0	−0.9	−11
MUE	0.8	3.7	16	1.1	3.7	23
Max	7.6	74	258	13	85	354
σ_{MSE}	1.6	12	44	2.6	13	57
$\tilde{N}_{sel}$	51	34	19	40	26	14
$\sigma_{\tilde{N}_{sel}}$	11	9	6	11	10	6
S.U.	7.7	26	140	15	55	380
			AS*n*			
MSE	0.2	1.5	−2.2	−0.1	−2.1	−6.5
MUE	0.6	6.7	33	0.6	4.6	18
Max	3.8	121	633	4.3	44	213
σ_{MSE}	1.1	20	99	1.1	11	41
$\tilde{N}_{sel}$	52	30	14	52	30	14
$\sigma_{\tilde{N}_{sel}}$	8	4	2	8	4	2
S.U.	7.2	37	370	7.1	37	350
			RS*n*			
MSE	0.3	0.2	0.9	0	0	-0.7
MUE	0.5	0.5	2.7	0.4	0.7	3.3
Max	6.9	3.2	44	3.8	5.2	32
σ_{MSE}	1.2	0.9	6.9	0.8	1.3	6.5
$\tilde{N}_{sel}$	73	53	33	68	47	28
$\sigma_{\tilde{N}_{sel}}$	13	12	9	16	13	9
S.U.	2.6	6.8	29	3.2	9.5	46

Abbreviations: MO, molecular orbital; MSE, mean signed error; MUE, mean unsigned error.

fact, only the AG5 set has one molecule with error greater than 10%, molecule **6**, with an $[\alpha]_D$ of 1.3 deg $(dm^{-1}\ (g/mL)^{-1})$ and an error of 0.4 deg $(dm^{-1}\ (g/mL)^{-1})$.

The same trends across selection criteria are observed with both basis sets, and $n = 4$ or even 3 are reasonable choices for the RS*n* criterion and aug-cc-pVTZ. Similar considerations apply to B3LYP/aug-cc-pVTZ, where the tight $n = 5$ choice provides both accurate results and large estimated speedup for all selection criteria, except for RS*n*, where smaller *n* values are also acceptable. For B3LYP, there are a few cases of errors above 10%, all related to small $[\alpha]_D$ values: three molecules with AG5 and one with RG5. The three compounds for AG5 are **6**, **18**, and **34**, where the errors are −12%, 19%, and 12%, respectively, from $[\alpha]_D = -4$, 1.5, and −9 deg $(dm^{-1}$

$(g/mL)^{-1}$), respectively; for RG5, molecule **34** has a 13% relative error corresponding to an absolute error of 1.26 deg (dm^{-1} $(g/mL)^{-1}$).

4 | DISCUSSION AND CONCLUSIONS

In this work, we show that accurate calculations of $[\alpha]_\omega$ can be performed with a subset of MOs in the solution of the CPKS equations, which determine the perturbed density used to evaluate the G' tensor. Selecting only the MOs for the excited configurations that contribute the most to the final value of the property can significantly reduce the computational cost of these calculations with minimal loss of accuracy. We propose two selection criteria for the relevant MOs, based on the guess density used in the iterative solution of the CPKS equations. With these criteria, we also propose two definitions of the selection threshold, an absolute value and a value relative to the largest absolute element in the guess, and various numerical values for the threshold. We compiled a test set of 51 chiral organic molecules with $[\alpha]_D$ values spanning five orders of magnitude, and we tested the various combinations of criteria and thresholds with two functionals, CAM-B3LYP and B3LYP, and two basis sets, aug-cc-pVDZ and aug-cc-pVTZ.

We find that all of the selection criteria work fairly well, as long as the threshold value n is large enough. The latter is the most important factor, as it correlates with the number of discarded MOs. The selected MOs typically are high energy occupied and low-energy virtual orbitals. This is because the working equations to solve the system in Equation (3) contain an energy denominator based on the energy difference between occupied and virtual MOs for a particular configuration.[16,32,33] Therefore, excited determinants built from MOs close to the Fermi energy provide larger contributions to the perturbed density. We find that $n = 5$ is a very effective choice of threshold in combination with the aug-cc-pVDZ basis set. These calculations are highly accurate, with low average errors (the largest MUE across the set is 0.7% and the largest MSE is −0.2%), small σ_{MSE} (the largest is 4.1%), very small errors overall (the maximum error across all 8 sets is 9.4%), and good speedup ranging from 2 to 4×. The $n = 4$ threshold provides much larger speedup, up to 11× faster than the full basis set. These calculations come with a somewhat worse accuracy than their $n = 5$ counterparts, as the largest MUE and MSE increase to 3.9% and −1.1%, respectively, and σ_{MSE} increases up to 8.5%. This threshold choice may be ideal for screening a large number of compounds, as the calculations can be performed quickly at a fairly high accuracy. The few large relative Max errors for the $n = 4$ calculations (see Table 1) are all due to small absolute values of $[\alpha]_D$. Furthermore, the corresponding absolute errors are at most 4 times smaller than the expected average error for CAM-B3LYP/aug-cc-pVDZ, and 3 times smaller than what is expected for B3LYP/aug-cc-pVDZ.[18,44] On the other hand, the $n = 3$ threshold is too loose, and it leads to low accuracy. With the larger basis set, aug-cc-pVTZ, the best threshold is $n = 5$, because the size of the basis set leads to a larger number of small density elements whose sum becomes important. Nonetheless, the accuracy of the $n = 5$ calculations with this basis set is better than the $n = 4$ equivalents with aug-cc-pVDZ, and the estimated speedup is comparable (7-8×). At a given n, our results indicate that a relative threshold is more accurate than an absolute one, and the G^2 criterion is more accurate than G^1. Thus, we recommend either the RG5 or AS5 criteria as the best compromise between computational cost and accuracy (the RS5 criterion would be more accurate but also more computationally intensive). We note that the analysis performed in this work is based on canonical MOs, which may or may not represent the most compact representation of the OR tensor.[28,45] It would be interesting to repeat this analysis using a localized MO basis, but the CPKS code in GAUSSIAN is only available in the canonical MO basis, thus beyond the scope of this work,

This proof-of-concept work shows that the selection of MOs prior to the solution of the CPKS equations, with careful choice of the selection criterion, can lead to large estimated speedup in calculations while accurately reproducing $[\alpha]_\omega$. These promising results now require an efficient implementation of the procedure for actual timing information. More importantly, we plan to extend this approach to CC methods, whose steeper scaling with system size compared to DFT may make the selection criteria even more beneficial for the fast evaluation of this molecular property.

ACKNOWLEDGMENT

We are grateful for the support from the National Science Foundation under Grant CHE-1650942.

ORCID

Marco Caricato https://orcid.org/0000-0001-7830-0562

REFERENCES

1. Vaccaro PH. *Comprehensive Chiroptical Spectroscopy, Instrumentation, Methodologies, and Theoretical Simulations*. 1 New York: John Wiley & Sons; 2012:265-324.
2. Polavarapu PL. Ab initio molecular optical rotations and absolute configurations. *Mol. Phys.* 1997;91:551-554.

3. Cheeseman JR, Frisch MJ, Devlin FJ, Stephens PJ. Hartree-Fock and density functional theory ab initio calculation of optical rotation using GIAOs: basis set dependence. *J. Phys. Chem. A*. 2000;104:1039-1046.
4. Grimme S. Calculation of frequency dependent optical rotation using density functional response theory. *Chem. Phys. Lett.* 2001;339:380-388.
5. Polavarapu PL. Optical rotation: recent advances in determining the absolute configuration. *Chirality*. 2002;14(10):768-781.
6. Autschbach J, Patchkovskii S, Ziegler T, van Gisbergen SJA, Jan BE. Chiroptical properties from time-dependent density functional theory. *II. Optical rotations of small to medium sized organic molecules. J. Chem. Phys.* 2002;117:581-592.
7. Ruud K, Helgaker T. Optical rotation studied by density-functional and coupled cluster methods. *Chem. Phys. Lett.* 2002;352:533-539.
8. Stephens PJ, McCann DM, Cheeseman JR, Frisch MJ. Determination of absolute configurations of chiral molecules using ab initio time-dependent density functional theory calculations of optical rotation: how reliable are absolute configurations obtained for molecules with small rotations? *Chirality*. 2005;17: 52-64.
9. Grimme S, Bahlmann A, Haufe G. Ab initio calculations for the optical rotations of conformationally flexible molecules: a case study on six-, seven-, and eight-membered fluorinated cycloalkanol esters. *Chirality*. 2002;14(10):793-797.
10. Ruud K, Stephens PJ, Devlin FJ, Taylor PR, Cheeseman JR, Frisch MJ. Coupled-cluster calculations of optical rotation. *Chem. Phys. Lett.* 2003;373:606-614.
11. Tam MC, Russ NJ, Crawford TD. Coupled cluster calculations of optical rotatory dispersion of (S)-methyloxirane. *J. Chem. Phys.* 2004;121(8):3550-3557.
12. Crawford TD. Ab initio calculation of molecular chiroptical properties. *Theor. Chem. Acc.* 2006;115:227-245.
13. Crawford TD, Stephens PJ. Comparison of time-dependent density-functional theory and coupled cluster theory for the calculation of the optical rotations of chiral molecules. *J. Phys. Chem. A*. 2008;112(6):1339-1345.
14. Pedersen TB. Koch H, Boman L, Sánchez De Merás AMJ. Origin invariant calculation of optical rotation without recourse to London orbitals. Chem. Phys. *Lett.* 2004;393:319-326.
15. Crawford TD, Owens LS, Tam MC, Schreiner PR, Koch H. Ab initio calculation of optical rotation in (P)-(+)-[4]triangulane. *J. Am. Chem. Soc.* 2005;127(5):1368-1369.
16. Krykunov M, Autschbach J. Calculation of optical rotation with time-periodic magnetic-field-dependent basis functions in approximate time-dependent density-functional theory. *J. Chem. Phys.* 2005;123:114103.
17. Autschbach J. Computing chiroptical properties with first-principles theoretical methods background and illustrative examples. *Chirality*. 2009;21:E116-E152.
18. Srebro M, Govind N, de Jong WA, Autschbach J. Optical rotation calculated with time-dependent density functional theory: the OR45 Benchmark. *J. Phys. Chem. A.* 2011;115(40):10930-10949.
19. Wiberg KB, Caricato M, Wang Y, Vaccaro PH. Towards the accurate and efficient calculation of optical rotatory dispersion using augmented minimal basis sets. *Chirality*. 2013;25(10): 606-616.
20. Baranowska-Łączkowska A, Łączkowski KZ. The ORP basis set designed for optical rotation calculations. *J. Comput. Chem.* 2013;34(23):2006-2013.
21. Baranowska-Łączkowska A, Łączkowski KZ, Henriksen C, Fernández B, Kozak M, Zielińska S. New basis set for the prediction of the specific rotation in flexible biological molecules. *RSC Adv.* 2016;6:19897-19902.
22. Pulay P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* 1983;100:151-154.
23. Saebø S, Pulay P. Local treatment of electron correlation. *Chem. Phys. Lett.* 1993;4:213-236.
24. Russ NJ, Crawford TD. Local correlation in coupled cluster calculations of molecular response properties. *Chem. Phys. Lett.* 2004;400:104-111.
25. Russ NJ, Crawford TD. Local correlation domains for coupled cluster theory: optical rotation and magnetic-field perturbations. *Phys. Chem. Chem. Phys.* 2008;10(23):3345-3352.
26. McAlexander HR, Mach TJ, Crawford TD. Localized optimized orbitals, coupled cluster theory, and chiroptical response properties. *Phys. Chem. Chem. Phys.* 2012;14(21):7830-7836.
27. McAlexander HR, Crawford TD. A comparison of three approaches to the reduced-scaling coupled cluster treatment of non-resonant molecular response properties. *J. Chem. Theory Comput.* 2016;12(1):209-222.
28. Aharon T, Caricato M. Configuration space analysis of the specific rotation of helicenes. *J. Phys. Chem. A.* 2019;123(20):4406-4418.
29. Caricato M. Orbital analysis of molecular optical activity based on configuration rotatory strength. *J. Chem. Theory Comput.* 2015;11(4):1349-1353.
30. Caricato M. Conformational effects on specific rotation: a theoretical study based on the $\tilde{S}_k$ method. *J. Phys. Chem. A.* 2015; 119:8303-8310.
31. Rosenfeld L. Quantenmechanische Theorie der natürlichen optischen Aktivität von Flüssigkeiten und Gasen. *Z. Phys.* 1929; 52:161-174.
32. Crawford TD, Tam MC, Abrams ML. The current state of ab initio calculations of optical rotation and electronic circular dichroism spectra. *J. Phys. Chem. A.* 2007;111(48):12057-12068.
33. Autschbach J. Time-dependent density functional theory for calculating origin-independent optical rotation and rotatory strength tensors. *ChemPhysChem.* 2011;12(17):3224-3235.
34. London F. Théorie quantique des courants interatomiques dans les combinaisons aromatiques. *J. Phys. Radium.* 1937;8:397-409.
35. Ditchfield R. Self-consistent perturbation theory of diamagnetism. *Mol. Phys.* 1974;27:789-807.
36. Pople J, Krishnan R, Schlegel B, Binkley J. Derivative studies in Hartree-Fock and Møller-Plesset theories. *Int. J. Quantum Chem.* 1979;16:225-241.
37. Frisch M, Head-Gordon M, Pople J. Direct analytic SCF second derivatives and electric field properties. *Chem. Phys.* 1990;141: 189-196.
38. Frisch MJ et al. *Gaussian Development Version*. Wallingford CT: Gaussian, Inc.; 2009.
39. Yanai T, Tew DP, Handy NC. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* 2004;393:51-57.

40. Woon DE, Dunning TH. Gaussian basis sets for use in correlated molecular calculations IV. *Calculation of static electrical response properties. J. Chem. Phys.* 1994;100:2975-2988.
41. Becke AD. A new mixing of Hartree-Fock and local density functional theories. *J. Chem. Phys.* 1993;98:1372-1377.
42. Becke AD. Density functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 1993;98:5648-5652.
43. Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B.* 1988;37:785-789.
44. Stephens PJ, Devlin FJ, Cheeseman JR, Frisch MJ. Calculation of optical rotation using density functional theory. *J. Phys. Chem. A.* 2001;105:5356-5371.
45. Moore B, Srebro M, Autschbach J. Analysis of Optical Activity in Terms of Bonds and Lone-Pairs: The Exceptionally Large Optical Rotation of Norbornenone. *J. Chem. Theory Comput.* 2012;8(11):4336-4346.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Aharon T, Caricato M. A molecular orbital selection approach for fast calculations of specific rotation with density functional theory. *Chirality*. 2020;32:243–253. https://doi.org/10.1002/chir.23158